DOCUMENT RESUME

ED 217 080 TM 820 342

AUTHOR Huntley, Renee M.; Loyd, Brenda H.
TITLE The Effect of Item Density on Item Difficulty in Passage-Related Language Tests.
PUB DATE Mar 82
NOTE 9p.; Paper presented at the Annual Meeting of the American Educational Research Association (66th, New York, NY, March 19-23, 1982).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Difficulty Level; *Language Tests; Secondary Education; *Test Construction; Test Format; *Test Items
IDENTIFIERS American College Testing Program; *Item Density

ABSTRACT

The study investigated the effect of "item density" on item difficulty in passage-related language tests. Item density refers to the number and frequency of items in relation to clear text. The format used was a passage with underlinings to signal the language situations out of which the items were constructed. American College Testing Program (ACT) Assessment experimental units were constructed that varied the number of items per sentence of text within the passage. Each of the 3 versions of the text contained 30 items and the proportional representation across the 5 item classifications was the same. Results suggest that item density significantly influences item difficulty. However, it could not be discerned which language classifications were particularly susceptible to the impact. The study suggested that the quality of passage-related language formats would be enhanced if test constructors presented no more than two test items per sentence. Achieving appropriate item density in these formats would seem to improve validity and reliability. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED217080

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

The Effect of Item Density on Item Difficulty in Passage-Related Language Tests

Renee M. Huntley

American College Testing Program

Brenda H. Loyd

University of Virginia

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. M. Huntley

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the annual meeting of the

American Educational Research Association,

New York, 1982

TM 820 342

The Effect of Item Density on Item Difficulty in Passage-Related Language Tests

RENEE M. HUNTLEY
American College Testing Program

BRENDA H. LOYD
University of Virginia

ABSTRACT

This study investigated the effect on item difficulty of "item density" -- that is, the number and frequency of items in relation to clear text -- in passage-related language tests. ACT Assessment experimental units were constructed that varied the number of items per sentence of text within the passage. Results suggest that item density significantly influences item difficulty, although which language classifications were particularly susceptible to the impact could not be discerned. Achieving appropriate item density in these formats would seem to improve validity and reliability.

The Effect of Item Density on Item Difficulty in Passage-Related Language Tests

RENEE M. HUNTLEY
American College Testing Program

BRENDA H. LOYD
University of Virginia

RENEE M. HUNTLEY
ACT
Test Development
P. O. Box 168
Iowa City, Iowa 52243

INTRODUCTION

In the assessment of language skills, test constructors have expressed a partiality for multiple-choice assessment instruments mainly because such tests do "an amazingly effective job of ordering students in the same way as a trained group of readers would after reading a sizable sample of their actual writing" (Godshalk, et al., p. 21). In addition, these measures can provide useful information about students' rhetorical abilities reliably and efficiently in terms of administration and money.

Among the various formats used in multiple-choice language tests is one that has the advantage of not violating the fundamental language tenet that language "situations" not be wrested from the context of "purpose, audience, intention, and occasion" (Huntley, p. 17). This format presents a continuous piece of discourse that has underlined and numbered portions. Alternative responses are provided, and the examinee must choose the one alternative that, for example, best expresses the idea, makes the statement appropriate for standard written English, or is worded most consistently with the style and tone of the passage as a whole.

Test constructors very often choose to construct a language test in this format precisely because it seems to address rhetorical concerns better than a format of discrete sentences might: without the context of an essay, how could one evaluate appropriateness, patterns, relationships of word choice to meaning, and so forth? But in developing such a test, what determines how many items should be drawn from each sentence of the essay? Although a test constructor could proceed from an intuitive sense of the amount of clear text needed between items so that examinees could at least follow the discourse, benefits would accrue to the examinees were research able to pinpoint the precise number of items per sentence that would improve the validity and reliability of the test. We would be able, for example, to determine whether the difficulty of the item was impacted by "item density" (the number and frequency of items in relation to clear text). And assurance would be provided that what was being tested was a construct (like punctuation, grammar, sentence structure, logic and organization), rather than some extraneous testing circumstance. Finally, we might be able to determine something about the nature of language were we to discover that one category of items (like sentence structure or diction and style, for example) was more susceptible to the impact of item density than another (like grammar). A recent literature search, however, indicates that this issue has not yet been investigated.

METHOD

This study investigated the effect of "item density" on item difficulty. Three versions of a language unit were prepared. The format used was that of a passage with underlinings to signal the language situations out of which the items were constructed. (Examples of these items can be found in the Appendix on page 6.) In Version A, thirty items were constructed across the essay, two items per sentence. The items were constructed to reflect a balanced weighting of classification -- that is, a proportional representation across five classifications. These classifications, which have been deemed crucial elements of effective expository prose, are punctuation, grammar, sentence structure, diction and style, and logic and organization.

Two additional versions were developed in the following manner: in Version B, nine items were selected from Version A and presented under the experimental density condition of one item per sentence. Thus, in Version B, these nine items appeared as the only test item in the sentence.

In Version C, another set of nine items from Version A was selected and presented under the experimental density condition of three items per sentence. Thus, in Version C, these nine items occurred as one of three test items in the sentence. It should be noted that some of the Version A items selected for Version C were the same as some selected for Version B. Although there had been an attempt made to select the same items as the focus of all the experimental units, the attempt could not be sustained. The problem seemed to arise in the multiple-item situation, in particular, and primarily because item development was completely dependent on the language possibilities inherent in the given sentence.

Each of the three versions of the test unit contained a total of thirty items, and the proportional representation across the five item classifications was the same for each test unit.

Each experimental test unit was administered to a random sample of approximately 250 high school students as part of the ACT Assessment tryout procedures.

ANALYSIS

The test items were scored as follows: a "1" was recorded for a correct response, and a "0" was recorded for an incorrect response. A two factor analysis of variance was used to determine the impact of item density on item difficulty. (Item difficulty is defined as the proportion of the population answering the item correctly. Thus, the higher the number that signifies item difficulty, the greater the proportion of the population answering correctly, and the easier the item.) The first factor was item density, and there were three levels of item density: one item per sentence of text, two items per sentence of text, and three items per sentence of text. The second factor was the selected sets of experimental items.

Two analyses were conducted. In the first analysis, the two density levels tested were the one-item-per-sentence level and the two-items-per-sentence level. In the second analysis, the two density levels tested were the two-items-per-

sentence level and the three-items-per-sentence level. In each analysis, the items factor consisted of the nine items common to the particular set of density conditions.

RESULTS

The first analysis of variance procedure indicated no significant difference in average item difficulty for test items that are presented one per sentence compared to their being presented as one of two items per sentence. The average item difficulty was .550 for the one-item-per-sentence density condition and .549 for the two-items-per-sentence density condition.

The second analysis of variance procedure indicated a significant effect for item density. The item difficulty of items that were embedded as one of two items per sentence was significantly higher (i.e., these items were easier in this presentation) than the item difficulty of these items when they were embedded as one of three items per sentence. The average item difficulty under the two-items-per-sentence density condition was .547; under the three-items-per-sentence density condition, it was .522. The difference was significant at a .06 level of significance.

To follow up the significant main density effect, the difference between the two items-per-sentence and three-items-per-sentence density levels was tested for each item individually. Four of these nine items tested showed statistically significant difference between the two density levels. Of these four items, two measured the construct of sentence structure, one measured diction and style, and one measured grammar.

DISCUSSION

The results of this study suggest that "item density" -- the number and frequency of items in relation to clear text in passage-related language tests -- significantly influences item difficulty under density conditions of three items per sentence. Less clear text and more items made for harder items. Less clear text that is more interrupted would adversely affect the examinee's ability to discern the meaning of the discourse -- in short, to make sense out of what was being read. This finding would support an intuitive or common sense approach to the construction of such test units, as would the finding of no significant difference in item difficulty under density conditions of two items per sentence. In this latter case, however, it might be fruitful to investigate whether it is the amount of clear text between items, rather than the number of items in the sentence, that is responsible for the effect.

Whether certain language classifications are particularly susceptible to the impact of item density cannot with certainty be determined. The four items in three item classifications so affected are simply an insufficient number from which to discern a clear trend, though they may suggest one that could be investigated in future studies. Rhetorical knowledge suggests that some language classifications would be likelier candidates for being affected than others -- situations, for example, where the examinee would have to see enough of a pattern to pick the appropriate option (as in sentence structure) or to perceive enough of the development of ideas to choose the appropriate relationship between them (as in logic and organization).

While future studies may help answer the concerns above, this study does suggest that the quality of passage-related language formats would be enhanced if the test constructor presented no more than two test items per sentence.

REFERENCES

Godshalk, F.I., Swineford, F., and Coffman, W.E. The Measurement of Writing Ability. New York: College Entrance Examination Board, 1966.

Huntley, Renee M. What's REALLY Basic in the Language Arts? Iowa City, Iowa: American College Testing Program, 1977.

APPENDIX

Sample Items

Version A:

One result of the growing interest in Africa in the past few years have been (1) an increased awareness of and appreciation for the African legacy given to (2) world culture....

2. A. NO CHANGE
B. which was given to
C. being given to
D. having been given to
E. to

(Item classification: Diction & Style)

Version B:

One result of the growing interest in Africa in the past few years has been an increased awareness of and appreciation for the African legacy given to (1) world culture....

1. Same item as above

Version C:

One result of the growing interest in Africa in the past few years have been (1) an increased awareness and appreciation for (2) the African legacy given to (3) world culture....

3. Same item as above

APPENDIX, Continued

Version A:

...African art typically uses vibrant colors, animated figures, geometrical forms, and materials that are natural (1) to emphasize the unity between human life and physical nature as a whole.... (2)

1. A. NO CHANGE
   B. natural materials
   C. materials which come from nature
   D. materials from nature
   E. materials of nature

(Item classification: Sentence Structure)

Version B:

...African art typically uses vibrant colors, animated figures, geometrical forms, and materials that are natural (1) to emphasize the unity between human life and physical nature....

1. Same item as above

Version C:

...African art typically uses vibrant colors, animated figures, geometrical forms, and materials that are natural (1) to emphasize the unity between human (2) life and physical nature as a whole.... (3)

1. Same item as above